Learning-based biases in quantity-insensitive stress

Robert Staubs

University of Massachusetts Amherst

Northeast Computational Phonology Circle V
October 15, 2010

It has often been noted that primary stress tends to fall on the "first" foot placed (e.g. Hayes 1995).

I propose an explanation of this tendency based in formal properties of the learning problem.

Structure of the talk:

1. Typological overview
2. Proposed explanation and model
3. Results
4. Extensions to other tendencies

Typological correlation

Languages which can be described as parsing first on a particular edge tend to place main stress on that same edge.

"First" as main stress in Heinz (2007):

- 14 of 14 types of quantity-insensitive iterative stress languages
- 68 of 68 of all quantity-insensitive iterative stress languages

Not categorically true, however.

All iterative stress languages (StressTyp/WALS; Goedemans and van der Hulst 2005):

	Left-to-right	Right-to-left
Left main	63	12
Right main	27	53

$\chi^2 = 38.1, p < 0.05$

This correlation is not absolute—counterexamples exist. Any explanation that requires compliance (e.g. van der Hulst 1984) is therefore untenable.

This correlation is not absolute—counterexamples exist. Any explanation that requires compliance (e.g. van der Hulst 1984) is therefore untenable.

The properties involved ("secondary stress" and "primary stress") are relatively abstract. Phonetics-based accounts of typological tendencies (e.g. Blevins 2004, Ohala 2005) seem unlikely.

This correlation is not absolute—counterexamples exist. Any explanation that requires compliance (e.g. van der Hulst 1984) is therefore untenable.

The properties involved ("secondary stress" and "primary stress") are relatively abstract. Phonetics-based accounts of typological tendencies (e.g. Blevins 2004, Ohala 2005) seem unlikely.

I place the explanation of this tendency in learning: a potentially probabilistic and "analogical" process.

A learning-based account

Same-side systems have a potential advantage in learning: secondary stress and main stress are mutually-reinforcing.

A learning-based account

Same-side systems have a potential advantage in learning: secondary stress and main stress are mutually-reinforcing.

A given learning datum potentially contains information about secondary stress locations and primary stress location.

A learning-based account

Same-side systems have a potential advantage in learning: secondary stress and main stress are mutually-reinforcing.

A given learning datum potentially contains information about secondary stress locations and primary stress location.

A reliable location for either could be learned more readily because of a greater degree of evidence.

A learning-based account

Same-side systems have a potential advantage in learning: secondary stress and main stress are mutually-reinforcing.

A given learning datum potentially contains information about secondary stress locations and primary stress location.

A reliable location for either could be learned more readily because of a greater degree of evidence.

If both locations are identical, the amount of evidence is increased—and so is learnability.

If biases are present in the learnability of patterns, these may be magnified in iterated learning (review in Kirby *et al.* 2007).

Thus the relative ease or difficulty in learning a pattern conceivably affects its frequency.

Moreton and Pater (2011) use this idea and a constraint-based framework to model feature economy effects in typology. I adopt a similar model.

Bane and Riggle (2008) also identify some formal correlates of QI stress pattern frequency. This work was a major influence on the results presented here.

Grammatical framework

For the tests presented, I used a Maximum Entropy grammar (Hayes and Wilson 2008) framework.

This framework is a weighted constraints framework like Harmonic Grammar (Legendre *et al.* 1990, Pater 2008), linking it with other phonological work.

It assigns probabilities to candidates, allowing languages to be partially (non-categorically) learned.

I assume constraints that are maximally simple while maintaining good typological coverage:

- Constraints match *n*-grams with *n* up to 4.
- The symbols allowed are word edges, unstressed syllables, secondary stresses, and primary stresses.
- Constraints on secondary stresses match primary stresses as well (but not vice versa).
- All *n*-grams present in the target data are used.

Similar to constraint induction of Hayes and Wilson (2008) and Pater (to appear).

Constraints Illustrated

Bigram constraint satisfaction for a form and its flipped main stress version:

Constraint	Satisfactions	
	#102020#	#202010#
#1	1	0
#2	1	1
10	1	1
20	3	3
01	0	1
02	2	2

Learning

1. A string from the teacher's (categorical) grammar is sampled according to an Exponential distribution on word length.
2. The learner (probabilistically) produces a stress pattern for that word length given its grammar.
3. If there is a mismatch, the learner updates its constraint weights

I use the Perceptron update rule: weights are updated by the scaled difference between the violations of the target and winner.

I allow constraints to take on positive and negative weights, though this is not needed for the material presented.

The degree to which a pattern is learned after a given number of iterations can be expressed by the difference between the categorical target distribution and the probabilistic learned distribution (averaged over a number of trials).

I use sum squared error here, though other metrics are possible (e.g. K-L divergence).

The learnability of different languages can be compared if the iteration of evaluation is held constant.

Here I report results form mean SSE at iteration 10,000 with learning rate 0.01.

Directional Biases

One way to measure the learning bias for same-edge stress is to compare the learnability of a pattern with a "flipped" version—one in which the edge of main stress is switched.

If we subtract the learning measure value for the flipped pattern from the starting one, we obtain a measure of bias:

- Difference negative: the original pattern is better-learned
- Difference zero: both patterns are equally well-learned
- Difference positive: the flipped pattern is better-learned

Schematic Bias Results

Foot Type	Direction	Binary	Learnability Difference
Trochees	Left-to-Right	yes	-0.826
		no	-1.071
	Right-to-Left	yes	-0.177
		no	0.580
Iambs	Left-to-Right	yes	-0.186
		no	0.570
	Right-to-Left	yes	-0.821
		no	-1.063

Merged Schematic Bias Results

Pattern	Learnability Difference
iterate from edge, perfect grid	-1.067
iterate from edge, lapse at opposite	-0.824
iterate from one off edge, perfect grid	-0.177
iterate from one off edge, clash at opposite	0.575

Pattern	Strings	Learnability Difference
from edge	102020 1020202	-1.067
from edge, lapse	102020 1020200	-0.824
from one off edge	010202 0102020	-0.177
from one off edge, clash	010202 0102022	0.575

Pattern	Strings	Learnability Difference
from edge	102020 1020202	-1.067
from edge, lapse	102020 1020200	-0.824
from one off edge	010202 0102020	-0.177
from one off edge, clash	010202 0102022	0.575

The only patterns with positive differences are those which would call for a clash with main stress.

These are reasonably likely to be marked on other grounds (Kager 2001).

Bias in Constraints

Patterns which place primary stress at the same spot as their first secondary stress are supported by multiple constraints.

Bigram constraint satisfaction for a form and its flipped main stress version:

Constraint	Satisfactions	
	#102020#	#202010#
#1	1	**0**
#2	1	**1**
10	1	1
20	3	3
01	0	1
02	2	2

Typological Results

Mean learnability differences for QI stress patterns in Heinz (2007):

All languages:	-0.242	(Min: -1.374,	Max: 1.058)
All single stress:	0.003	(Min: -0.023,	Max: 0.039)

Typological Results

Mean learnability differences for QI stress patterns in Heinz (2007):

All languages:	-0.242	(Min: -1.374,	Max: 1.058)
All single stress:	0.003	(Min: -0.023,	Max: 0.039)
Main stress same:	-0.271	(Min: -1.374,	Max: 1.058)
Main stress opposite:	-0.664	(Min: -1.081,	Max: -0.246)

Here "same" means languages left for left-to-right parsing and right for right-to-left parsing.

Patterns with main stress opposite are markedly “non-reversible.”

The non-flipped/flipped difference is -0.664.

These two languages are bidirectional trochaic systems.

These languages do seem to place main stress this way (Alber 2005).

We see that having an isolated secondary stress foot on the opposite side of the word would be less learnable than a primary stress foot.

The learnability account thus does not just predict a simple bias towards “same side” stress.

Instead it supports the view of placing primary stress on the “first” foot.

Patterns with main stress on the same side as the start of parsing are mixed.

Most are non-reversible: without the four languages with positive scores, the mean value is -0.737.

The reversible languages are the following:

1. Left-to-right trochees with final stress (and degenerate feet)
2. Right-to-left iambs with initial stress (and degenerate feet)
3. Left-to-right iambs with penultimate stress
4. Right-to-left trochees with initial stress

1.	102022 1020202
2.	220201 2020201
3.	010220 0102020
4.	202010 2202010

Each of these languages clashes when parsing meets the fixed stress.

As before, there should be some extra pressure against clashes at peaks.

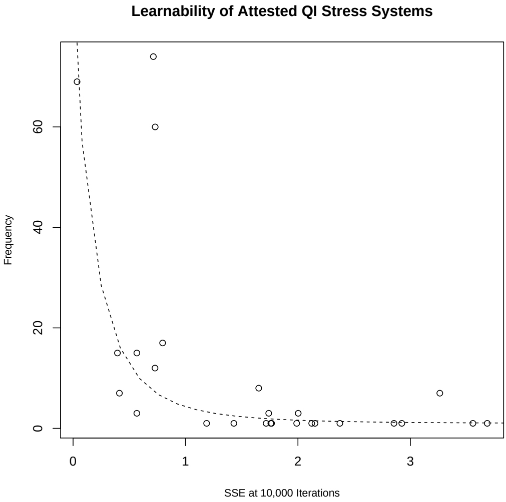
To summarize:

- In first-foot primary stress systems, the relevant stress reliably matches constraints on both primary and secondary stress.
- This duplication of evidence leads to faster learning in the model presented.
- This generally leads to a bias for same-edge stress—but for opposite-edge stress in some bidirectional systems (in which this does seem to be normal).

Other Implications

This model predicts other facts on the basis of learning:

- Single stress systems should be close to the word edge.
- Binary systems should be more common than ternary ones.
- Single stress systems should be more common than iterative ones.
- ...



Exponential regression, $r^2 = 0.525$

Conclusion

I described an approach to typological tendencies in stress systems based on their learnability in a particular learning model.

This learning model is general-purpose and already somewhat established in the field.

Generally, this work supports a way of thinking about typological tendencies as emerging from the formal properties of the patterns, rather than any *a priori* assertions about grammar.

Thanks are owed especially to Jeff Heinz for his typological database and Max Bane for his processing of it.

This research was supported by the National Science Foundation under grant number BCS-0813829 (2008-2012) to the University of Massachusetts, Amherst, and by a National Science Foundation Graduate Research Fellowship to the author.

Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.